

Ali Janati

aj3238@columbia.edu | (510) 423-2697 | [linkedin.com/in/alijanati/](https://www.linkedin.com/in/alijanati/) | github.com/Na00s | alijanati.netlify.app

Columbia MSc. In Data Science Candidate

EDUCATION

Columbia University

Aug 2023 - Dec 2024

MSc. in Data Science

New York, NY

- Relevant coursework: Algorithms for Data Science, Applied Machine Learning, Applied Deep Learning, Deep Learning for NLP.
- Research Project: Working under Prof. Pierre Gentine as a Graduate Research assistant to fine-tune IBM - NASA geospatial Vision Transformer (ViT) foundation model for recognizing wind damaged forest areas through image segmentation.

Mines Paris

Sep 2018 - Jun 2023

MSc. and BSc. in Engineering and Computer Science

Paris, France

- French Top-ranking Master & Bachelor of Science, "Grande Ecole" program. Joint Master of Engineering with **ESPCI Paris - PSL**.
- Relevant coursework: Applied Mathematics, Machine Learning, Applied statistics, Computer Science, Fundamental Physics.

Lycee Lakanal

Aug 2015 - Aug 2018

BSc. in Applied Physics

Paris, France

- Undergraduate program preparing for the highly competitive nationwide entrance exams to the Top-ranking French Schools of Engineering. (CPGE PCSI/ PC*).
- Relevant coursework: Mathematics, Physics, Chemistry, Computer Science, Linear Algebra, Calculus.

WORK EXPERIENCE

Esperanto Technologies

Jun 2024 - Present

Machine Learning Engineer Intern

Mountain View, CA

- Operating at Esperanto, a tech startup developing AI/ML semiconductors to speed inference using RISC-V architected chips.
- Benchmarked and Fine-tuned different versions of the OpenAI Whisper speech recognition model through transfer learning on domain specific datasets to enhance performance on medical conversations between doctors and patients. Achieved 14% WER drop.
- Currently identifying layers that can be deleted in Llama 3.1 through several pruning techniques while keeping the highest MMLU Accuracy possible, healing the pruned model through PEFT to recover any lost performance by the pruning process.

Ekimetrics

Nov 2022 - Apr 2023

Machine Learning Engineer Intern

Paris, France

- Engaged in Ekimetrics, a pioneering leader in data science and AI-powered solutions for industry.
- Built an NLP model to monitor market and competitor activity by identifying companies with positive mentions.
- Boosted existing models, achieving a 14% increase in f1 score by implementing and benchmarking advanced techniques ranging from random forest to transformer encoders.

Humanitics.ai, Station F

Apr 2022 - Oct 2022

Machine Learning Engineer Intern

Paris, France

- Contributed to Humanitics.ai, a retail-tech within world's largest startup campus: station F, and laureate of future 40 most promising startups in France in 2022.
- Shaped a traffic forecasting model to determine for each one of 500 stores, and each hour of the day in the future, the number of visitors leveraging sophisticated naïve models, prophet and PyTorch and exceeded performance expectations by 7% MAPE drop.

CNRS (French National Centre for Scientific Research)

May 2021 - Aug 2021

Research Assistant in Computer Science

Paris, France

- Spearheaded developments of python-based simulation models, enhancing acoustic imagery through doppler effects, time accelerated preliminary research outcomes by 15%.
- Conducted extensive bibliographic research, implemented tailored models for 5 experiments, and analyzed data to extract key insights for researchers on the feasibility of said experiments.

ESPCI & PSL University

Jan 2021 - Apr 2021

Research Assistant in Deep Learning, Part-time

Paris, France

- Operated in ESPCI & PSL research centers as a research assistant to classify images of bidimensional spin models according to whether the total topological charge is null or not.
- Developed and optimized multiple deep neural network architectures using PyTorch, achieving 93% recall in classifying bidimensional spin models, thereby enhancing predictive reliability and contributing to innovative research methodologies.

PUBLICATIONS

- Dumont, E. L. P., Janati, A., Bhattacharya, M., Jeannin, J.-B., Do, C. (2024). Improving allele-specific epigenomic signal coverage by 10-fold using Hidden Markov Modeling and Machine Learning. bioRxiv, 2024.05.23.595536. <https://doi.org/10.1101/2024.05.23.595536>

PROJECTS

Music Labs

- Developing a music centered application containing an Image to Music feature with a personalized version of Hugging Face hosted llava-v1.5-13b and a Sentence to Playlist recommendation feature using TFIDF.

Mini GPT

- Training a Transformer Decoder, inspired from the "Attention is all you need paper", on Machine learning lecture slides to generate meaningful character level tokens.

SKILLS

Programming: Fluent in Python (Numpy, Pandas, ScikitLearn, PyTorch, Plotly, Hugging Face), Git, Google Cloud.

Languages: English (fluent, TOEFL score: 109/120). French and Arabic (native speaker). Proficient in Spanish.

Associative Work: Tutoring for preparatory classes (CPGE) and high school students, Data for Good.

Sports: Soccer (played in Raja Casablanca at a national level from 2009 to 2014), Fitness, Martial Arts.